1

Denote the proportion of people aged over 55 who are willing to take a calss as p_1, the proportion of people aged 18-55 who are willing to take a calss as p_2.

$H_0 : p_1 {=} p_2$ $H_1 : p_1 \neq p_2$

$$\hat{p_1} = 51/170 = 0.3$$

$$\hat{p_2} = 79/230 = 0.3435$$

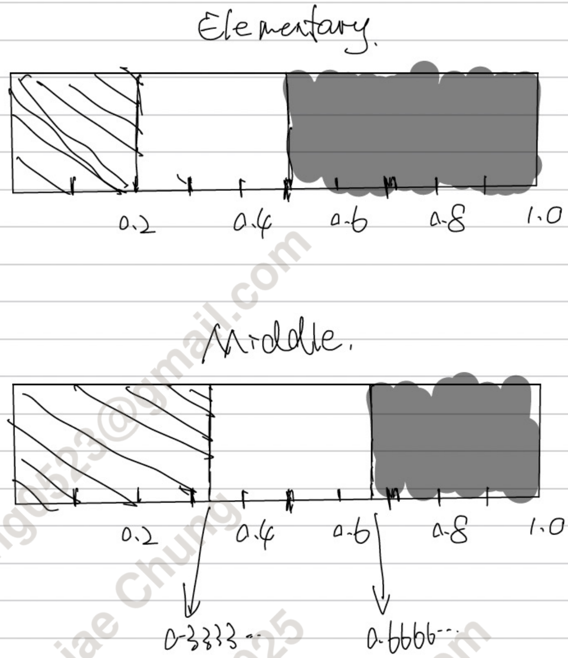
$$Z = \frac{\hat{p_2} - \hat{p_1}}{\sqrt{\frac{\hat{p_2}(1-\hat{p_2})}{n_2} + \frac{\hat{p_1}(1-\hat{p_1})}{n_1}}} = 0.9241$$

Since significance level $\alpha = 0.05$, and by the table we know $Z_{0.025} = 1.96 > 0.9241$. Thus, we cannot reject null hypothesis H_0, which means that there's no difference between two proportions.

2

Q2.

(a)



对于Middle School, 三等分 ☆

(b) Assume that the number of bottles sold by elementary school students is n, then the number of small bottles is 0.5n. Therefore, the number of bottles sold by middle school students is 3n, where the number of small bottles is n. Since n¿0.5n, the conclusion by the elementary administrator is wrong.
(c) i For school A, the proportion of small bottles ito medium to large is: 0.15: 0.15:0.7.

For school B, the proportion of small bottles ito medium to large is: 0.2: 0.2:0.6. So the school B sold a greater proportion of large bottles.

(c) ii Let the total number of bottles sold by school A be N_a, the total number of bottles sold by school A be N_b. Thus the number of large bottles sold by school A is $0.7N_a$, the number of large bottles sold by school A is $0.6N_a$

Let $0.7N_a > 0.6N_a$, we get $\frac{N_a}{N_b} > \frac{7}{6}$.

So when $\frac{N_a}{N_b} > \frac{7}{6}$, school A sold more large bottles. When $\frac{N_a}{N_b} = \frac{7}{6}$, school A and B sold same large bottles. When $\frac{N_a}{N_b} < \frac{7}{6}$, school B sold more large bottles.

3

(a) This was an observational study as there was no human intervention by the researcher to form the data and all data was obtained simply by contacting others.

(b) Using completely random sampling, the 70 days were numbered 1-70, from which 35 numbers were randomly selected. Next, autopilot was used on the selected days and not on the unselected days.

(c) Randomly select 70 model D cars and randomly divide these 70 model D into two groups of 35 cars each. The first group of cars all use autopilot to get to work, and the second group of cars do not use autopilot to get to work. The results obtained in this way can be generalized to all model D in the club.

4.

(a) Let X be the number of crystals opened until red crystal is found. So X is the geometric distribution,

$X \sim Ge(0.08)$

$$P(X=k) = (1-p)^{k-1} p \quad , \; p = 0.08$$

$$\Rightarrow E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$

$$= p \cdot \frac{d \sum_{k=1}^{\infty} (1-p)^k}{d(1-p)}$$

$$= p \cdot \frac{d \frac{1}{1-(1-p)}}{d(1-p)}$$

$$= p \cdot \frac{1}{(1-(1-p))^2}$$

$$= \frac{1}{p} = 12.5$$

ii $E(X^2) = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1}$

$$= p\left[\sum_{k=1}^{\infty} k(k-1)p(1-p)^{k-1} + \sum_{k=1}^{\infty} k(1-p)^{k-1}\right]$$

$$= p(1-p)\sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} + \frac{1}{p}$$

$$= p(1-p)\cdot\frac{2}{(1-(1-p))^3} + \frac{1}{p}$$

$$= \frac{2(1-p)}{p^2} + \frac{p}{p^2}$$

$$= \frac{2-p}{p^2}$$

$SD(X) = \sqrt{E(X^2) - E^2(X)} = \sqrt{\frac{2-p}{p^2} - \frac{1}{p^2}}$

$$= \sqrt{\frac{1-p}{p^2}}$$

$$= \sqrt{\frac{0.92}{0.08^2}}$$

$= 11.9896.$

(b) i.

$P(Y=3) = (1-p)^2 p = 0.92^2 \times 0.08 = 0.067712$

ii

$P(Y=4) = 1 - P(Y=1) - P(Y=2) - P(Y=3)$

$= 0.778688$

(c) i

$E(Y) = 1 \times 0.08 + 2 \times 0.0736 + 3 \times 0.067712$

$+ 4 \times 0.778688$

$= 3.545088$

ii

The average number of geodes open when Conrad stops playing is 3.545088.

5

(a) $p = \frac{71+76+112}{500} = 0.518$

(b) $p = \frac{80}{91}$

(c) i The chi-square independence test should be used

(c) ii H0:The number of months to collect baseball cards is independent of the type of cards collected

H1:The number of months to collect baseball cards is not independent of the type of cards collected.

(c) iii Since $p = 0.0075 < \alpha = 0.05$. So the null hypothesis is rejected, there is a relationship between the number of months of collecting baseball cards and the type of cards collected.

6

(a) i He should get the prices and sales of such whistles in these stores and calculate the whistle's weighted average price.

(a) ii Suppose the sales of whistles at store i are n_i and the price of whistles at store i is P_i,then the mean

$$\overline{P} = \frac{\sum n_i Pi}{\sum n_i}$$

(b) i Since the average number of the whistle is 5.12, and the medium number of the whistle is 4.885, the average number is higher than the medium number. So the data is right skewed distribution.

(b) ii

$$IQR = Q_3 - Q_1 = 5.475 - 4.51 = 0.965$$

$$Q_1 - 1.5IQR = 4.51 - 1.5 * 0.965 = 3.0625$$

$$Q_3 + 1.5IQR = 5.475 + 1.5 * 0.965 = 6.9225$$

Since the maximum value is 6.58 less than 6.9225 and the minimum value is 4.25 greater than 3.0625, there are no outliers.

(c) i

$$\frac{3(5.12 - 4.885)}{0.743} = 0.9489$$

(c) ii Since the skewness coefficient is 0.0489, the whistle price is right-skewed.

(d) Julia's data does not satisfy the normality condition because the sample is less than 30 and is right-skewed.